

WHAT IS CLAIMED IS:

1. A high-throughput method for determining a biochemical function of a protein or polypeptide domain of unknown function comprising:
 - (A) identifying a putative polypeptide domain that properly folds into a stable polypeptide domain, said stable polypeptide having a defined three dimensional structure;
 - (B) determining three dimensional structure of the stable polypeptide domain from an automated analysis of NMR spectrometer spectra of said polypeptide domain, wherein said automated analysis is conducted by a NOESY_Assign process;
 - (C) comparing the determined three dimensional structure of the stable polypeptide domain to known three-dimensional structures in a protein data bank, wherein said comparison identifies known structures within said protein data bank that are homologous to the determined three dimensional structure; and
 - (D) correlating a biochemical function corresponding to the identified homologous structure to a biochemical function for the stable polypeptide domain.
2. The method according to claim 1, further comprising the prestep of parsing a target polynucleotide into at least one putative polypeptide domain.
3. The method according to claim 2, wherein said parsing is performed by a first computer algorithm, wherein said first computer algorithm is selected from the group consisting of a computer algorithm capable of determining exon phase boundaries of a polynucleotide, and a computer algorithm capable of determining interdomain boundaries encoded in a polynucleotide.
4. The method of claim 3, further comprising a computer algorithm that compares the putative polypeptide domain sequence with known domain sequences stored within a database.
5. The method of claim 1, wherein said NMR spectra are analyzed by a second computer algorithm that automatically assigns resonance assignments to the polypeptide sequence.

PCT/US99/16417

-46-

6. The method of ~~claim 1~~, wherein said identification of said stable polypeptide domain comprises measuring a time course of amide hydrogen-deuterium exchange.

7. The method of ~~claim 1~~, wherein prior to step (B), said stable polypeptide domain is optimally solubilized, said optimum solubilization comprising:

- preparing an array of microdialysis buttons, wherein each of said microdialysis buttons contains at least 1 μ l of an approximately 1mM solution of said stable polypeptide domain;
- dialyzing each member of said array of microdialysis buttons against a different dialysis buffer;
- analyzing each of said dialyzed microdialysis buttons to determine whether said stable polypeptide domain has remained soluble; and
- selecting the polypeptide domain having optimum solubility characteristics for NMR spectroscopy.

15 8. The method of ~~claim 1~~, wherein said comparison of said determined three dimensional structure to said known three-dimensional structures in the protein data bank is performed by a third computer algorithm that is capable of determining 3D structure homology between said determined three dimensional structure and a member of said PDB.

20 9. The method according to ~~claim 11~~, wherein said third computer algorithm is selected from the group consisting of DALI, CATH and VAST.

10. The method of ~~claim 1~~, wherein said protein data bank is Protein Data Base ("PDB").

25 11. The method of ~~claim 4~~, wherein said database contains domain sequence information of known and determined domain sequences.

12. An integrated system for rapid determination of a biochemical function of a protein or protein domain of unknown function:

- a first computer algorithm capable of parsing said target polynucleotide into at least one putative domain encoding region;
- 30 (B) a designated lab for expressing said putative domain;

*SUB
B1*

(C) an NMR spectrometer for determining individual spin resonances of amino acids of said putative domain;

(D) a data collection device capable of collecting NMR spectral data, wherein said data collection device is operatively coupled to said NMR spectrometer;

5 (E) at least one computer;

(F) a second computer algorithm capable of assigning individual spin resonances to individual amino acids of a polypeptide;

(G) a third computer algorithm capable of determining tertiary structure of a polypeptide, wherein said polypeptide has had resonances assigned to individual amino acids of said polypeptide;

10 (H) a database, wherein stored within said database is information about the structure and function of known proteins and determined proteins; and

(I) a fourth computer algorithm capable of determining 3D structure homology between the determined three-dimensional structure of a polypeptide of unknown function to three-dimensional structure of a protein of known function, wherein said protein of known structure is stored within said protein database, wherein said fourth computer algorithm determines said structure by an automated NOESY_Assign process.

15

20

13. A high-throughput method for determining a biochemical function of a polypeptide of unknown function encoded by a target polynucleotide comprising the steps:

(A) identifying at least one putative polypeptide domain encoding region of the target polynucleotide ("parsing");

25 (B) expressing said putative polypeptide domain;

(C) determining whether said expressed putative polypeptide domain forms a stable polypeptide domain having a defined three dimensional structure ("trapping");

30 (D) determining the three dimensional structure of the stable polypeptide domain by an automated NOESY_Assign process;

(E) comparing the determined three dimensional structure of the stable polypeptide domain to known three dimensional structures in a Protein

-48-

Data Bank to determine whether any such known structures are homologous to the determined structure; and

(F) correlating a biochemical function corresponding to the homologous structure to a biochemical function for the stable polypeptide domain.

000744000-030201